# Glass Identification Data Set

*Download*: Data Folder, Data Set Description

**Abstract**: From USA Forensic Science Service; 6 types of glass; defined in terms of their oxide content (i.e. Na, Fe, K, etc)

| | | | | | |
|---|---|---|---|---|---|
| **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 214 | **Area:** | Physical |
| **Attribute Characteristics:** | Real | **Number of Attributes:** | 10 | **Date Donated** | 1987-09-01 |
| **Associated Tasks:** | Classification | **Missing Values?** | No | **Number of Web Hits:** | 476877 |

## Source:

Creator:

B. German
Central Research Establishment
Home Office Forensic Science Service
Aldermaston, Reading, Berkshire RG7 4PN

Donor:

Vina Spiehler, Ph.D., DABFT
Diagnostic Products Corporation
(213) 776-0180 (ext 3014)

## Data Set Information:

Vina conducted a comparison test of her rule-based system, BEAGLE, the nearest-neighbor algorithm, and discriminant analysis. BEAGLE is a product available through VRS Consulting, Inc.; 4676 Admiralty Way, Suite 206; Marina Del Ray, CA 90292 (213) 827-7890 and FAX: -3189. In determining whether the glass was a type of "float" glass or not, the following results were obtained (# incorrect answers):

Type of Sample -- Beagle -- NN -- DA
Windows that were float processed (87) -- 10 -- 12 -- 21
Windows that were not: (76) -- 19 -- 16 -- 22

The study of classification of types of glass was motivated by criminological investigation. At the scene of the crime, the glass left can be used as evidence...if it is correctly identified!

## Attribute Information:

1. Id number: 1 to 214
2. RI: refractive index
3. Na: Sodium (unit measurement: weight percent in corresponding oxide, as are attributes 4-10)
4. Mg: Magnesium
5. Al: Aluminum
6. Si: Silicon
7. K: Potassium
8. Ca: Calcium
9. Ba: Barium
10. Fe: Iron

11. Type of glass: (class attribute)
-- 1 building_windows_float_processed
-- 2 building_windows_non_float_processed
-- 3 vehicle_windows_float_processed
-- 4 vehicle_windows_non_float_processed (none in this database)
-- 5 containers
-- 6 tableware
-- 7 headlamps

### Relevant Papers:

Ian W. Evett and Ernest J. Spiehler. Rule Induction in Forensic Science. Central Research Establishment. Home Office Forensic Science Service. Aldermaston, Reading, Berkshire RG7 4PN
[Web Link]

### Papers That Cite This Data Set[1]:


Ping Zhong and Masao Fukushima. A Regularized Nonsmooth Newton Method for Multi-class Support Vector Machines. 2005. [View Context].

Vassilis Athitsos and Stan Sclaroff. Boosting Nearest Neighbor Classifiers for Multiclass Recognition. Boston University Computer Science Tech. Report No, 2004-006. 2004. [View Context].

Yuan Jiang and Zhi-Hua Zhou. Editing Training Data for kNN Classifiers with Neural Network Ensemble. ISNN (1). 2004. [View Context].

S. Augustine Su and Jennifer G. Dy. Automated hierarchical mixtures of probabilistic principal component analyzers. ICML. 2004. [View Context].

Xiaoli Z. Fern and Carla Brodley. Solving cluster ensemble problems by bipartite graph partitioning. ICML. 2004. [View Context].

Francesco Masulli. An experimental analysis of the dependence among codeword bit errors in ECOC learning machines. and Giorgio Valentini b,c. 2003. [View Context].

Krzysztof Krawiec. Genetic Programming-based Construction of Features for Machine Learning and Knowledge Discovery Tasks. Institute of Computing Science, Poznan University of Technology. 2002. [View Context].

Michail Vlachos and Carlotta Domeniconi and Dimitrios Gunopulos and George Kollios and Nick Koudas. Non-linear dimensionality reduction techniques for classification and visualization. KDD. 2002. [View Context].

Giorgio Valentini and Francesco Masulli. NEURObjects: an object-oriented library for neural network development. Neurocomputing, 48. 2002. [View Context].

D. I. S I and Francesco Masulli and Giorgio Valentini and D. I. S. Universit#a di Genova. Dipartimento di Informatica e Scienze dell' Informazione. 2001. [View Context].

Carlotta Domeniconi and Jing Peng and Dimitrios Gunopulos. An Adaptive Metric Machine for Pattern Classification. NIPS. 2000. [View Context].

Mark A. Hall. Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning. ICML. 2000. [View Context].

Petri Kontkanen and Petri Myllym and Tomi Silander and Henry Tirri and Peter Gr. On predictive distributions and Bayesian networks. Department of Computer Science, Stanford University. 2000. [View Context].

Thierry Denoeux. A neural network classifier based on Dempster-Shafer theory. IEEE Transactions on Systems, Man, and Cybernetics, Part A, 30. 2000. [View Context].

Francesco Masulli and Giorgio Valentini. Effectiveness of Error Correcting Output Codes in Multiclass Learning Problems. Multiple Classifier Systems. 2000. [View Context].

Nir Friedman and Iftach Nachman. Gaussian Process Networks. UAI. 2000. [View Context].

Kai Ming Ting and Ian H. Witten. Issues in Stacked Generalization. J. Artif. Intell. Res. (JAIR, 10. 1999. [View Context].

Christopher J. Merz. Using Correspondence Analysis to Combine Classifiers. Machine Learning, 36. 1999. [View Context].

Eibe Frank and Ian H. Witten. Generating Accurate Rule Sets Without Global Optimization. ICML. 1998. [View Context].

Jan C. Bioch and D. Meer and Rob Potharst. Bivariate Decision Trees. PKDD. 1997. [View Context].

D. Greig and Hava T. Siegelmann and Michael Zibulevsky. A New Class of Sigmoid Activation Functions That Don't Saturate. 1997. [View Context].

Christopher J. Merz. Combining Classifiers Using Correspondence Analysis. NIPS. 1997. [View Context].

. Prototype Selection for Composite Nearest Neighbor Classifiers. Department of Computer Science University of Massachusetts. 1997. [View Context].

Georg Thimm and E. Fiesler. Optimal Setting of Weights, Learning Rate, and Gain. E S E A R C H R E P R O R T I D I A P. 1997. [View Context].

Richard Maclin and David W. Opitz. An Empirical Evaluation of Bagging and Boosting. AAAI/IAAI. 1997. [View Context].

Ethem Alpaydin. Voting over Multiple Condensed Nearest Neighbors. Artif. Intell. Rev, 11. 1997. [View Context].

Ron Kohavi and Mehran Sahami. Error-Based and Entropy-Based Discretization of Continuous Features. KDD. 1996. [View Context].

Aynur Akkus and H. Altay Güvenir. K Nearest Neighbor Classification on Feature Projections. ICML. 1996. [View Context].

Jitender S. Deogun and Vijay V. Raghavan and Hayri Sever. Exploiting Upper Approximation in the Rough Set Methodology. KDD. 1995. [View Context].

Thomas G. Dietterich and Ghulum Bakiri. Solving Multiclass Learning Problems via Error-Correcting Output Codes. CoRR, csAI/9501101. 1995. [View Context].

. Ectiveness of Error Correcting Output Coding methods in ensemble and monolithic learning machines. Dipartimento di Informatica, Universitdi Pisa. [View Context].

Zhi-Hua Zhou and Xu-Ying Liu. Training Cost-Sensitive Neural Networks with Methods Addressing the Class Imbalance Problem. [View Context].

Aynur Akku and H. Altay Guvenir. Weighting Features in k Nearest Neighbor Classification on Feature Projections. Department of Computer Engineering and Information Science Bilkent University. [View Context].

Francesco Masulli and Giorgio Valentini. Quantitative Evaluation of Dependence among Outputs in ECOC Classifiers Using Mutual Information Based Measures. Universitdi Genova DISI - Dipartimento di Informatica e Scienze dell'Informazione INFM - Istituto Nazionale per la Fisica della Materia. [View Context].

Rong-En Fan and P. -H Chen and C. -J Lin. Working Set Selection Using the Second Order Information for Training SVM. Department of Computer Science and Information Engineering National Taiwan University. [View Context].

Yin Zhang and W. Nick Street. Bagging with Adaptive Costs. Management Sciences Department University of Iowa Iowa City. [View Context].

Ping Zhong and Masao Fukushima. Second Order Cone Programming Formulations for Robust Multi-class Classification. [View Context].

Karthik Ramakrishnan. UNIVERSITY OF MINNESOTA. [View Context].

Pramod Viswanath and M. Narasimha Murty and Shalabh Bhatnagar. A pattern synthesis technique to reduce the curse of dimensionality effect. E-mail. [View Context].

Erin J. Bredensteiner and Kristin P. Bennett. Multicategory Classification by Support Vector Machines. Department of Mathematics University of Evansville. [View Context].

Pramod Viswanath and M. Narasimha Murty and Shalabh Bhatnagar. Partition Based Pattern Synthesis Technique with Efficient Algorithms for Nearest Neighbor Classification. Department of Computer Science and Automation, Indian Institute of Science. [View Context].

Federico Divina and Elena Marchiori. Handling Continuous Attributes in an Evolutionary Inductive Learner. Department of Computer Science Vrije Universiteit. [View Context].

James J. Liu and James Tin and Yau Kwok. An Extended Genetic Rule Induction Algorithm. Department of Computer Science Wuhan University. [View Context].

Francesco Masulli and Giorgio Valentini. Comparing Decomposition Methods for Classification. Istituto Nazionale per la Fisica della Materia DISI - Dipartimento di Informatica e Scienze dell'Informazione. [View Context].

Alexander K. Seewald. Dissertation Towards Understanding Stacking Studies of a General Ensemble Learning Scheme ausgefuhrt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Naturwissenschaften. [View Context].

H. Altay G uvenir and Aynur Akkus. WEIGHTED K NEAREST NEIGHBOR CLASSIFICATION ON FEATURE PROJECTIONS. Department of Computer Engineering and Information Science Bilkent University. [View Context].

Ron Kohavi and Brian Frasca. Useful Feature Subsets and Rough Set Reducts. the Third International Workshop on Rough Sets and Soft Computing. [View Context].

H. Altay Guvenir. A Classification Learning Algorithm Robust to Irrelevant Features. Bilkent University, Department of Computer Engineering and Information Science. [View Context].

Stefan Aeberhard and Danny Coomans and De Vel. THE PERFORMANCE OF STATISTICAL PATTERN RECOGNITION METHODS IN HIGH DIMENSIONAL SETTINGS. James Cook University. [View Context].

Suresh K. Choubey and Jitender S. Deogun and Vijay V. Raghavan and Hayri Sever. A comparison of feature selection algorithms in the context of rough classifiers. [View Context].

Chih-Wei Hsu and Cheng-Ru Lin. A Comparison of Methods for Multi-class Support Vector Machines. Department of Computer Science and Information Engineering National Taiwan University. [View Context].

C. Titus Brown and Harry W. Bullen and Sean P. Kelly and Robert K. Xiao and Steven G. Satterfield and John G. Hagedorn and Judith E. Devaney. Visualization and Data Mining in an 3D Immersive Environment: Summer Project 2003. [View Context].

## Citation Request:

Please refer to the Machine Learning Repository's citation policy

---

[1] Papers were automatically harvested and associated with this data set, in collaboration with Rexa.info